January 2014

A Statistical Analysis of the Effects of Intrapartum Medications on Newborns in the First Hour After Birth

Jocelyn E. Pitman
Worcester Polytechnic Institute

Project Number: SW1301

A Statistical Analysis of the Effects of Intrapartum Medications on Newborns in the First Hour After Birth

A Major Qualifying Project Report

submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Jocelyn Pitman

Date: January 30 2014

Approved:

Professor Suzanne L. Weekes, Advisor

Abstract

Intrapartum medications, medications administered while a mother is in labor, are widely used and their role in delivery is not often questioned. The focus of this project is to examine the effects of a mother's intrapartum medications on the behaviors of a baby in the first hour after birth. Data was collected from 67 mothers during labor and in the first hour following labor in a hospital setting. We aim to use statistical tests to determine if there are relationships between the medications mothers receive and the behaviors a baby exhibits in the first hour after birth. There has been little research done in this area, so we hope to encourage further study.

Acknowledgments

First and foremost I would like to thank Professor Suzanne Weekes for her help in advising this project. I would like to thank Dr. Kajsa Brimdyr and Judy Blatchford at the Healthy Children Project for their tireless commitment to this study and for gathering the data that made this project possible. I would also like to thank Nancy Howard and Joshua Brimdyr for their encouragement and for being my test readers throughout the course of this project.

Contents

Chapter 1

Introduction

The following is a Major Qualifying Project created in conjunction with a study currently being conducted by The Healthy Children Project, Inc [4]. The working title of the study is "Effects of Intrapartum Analgesia and Anesthesia on Newborn Behaviors while Skin to Skin with Mother in the First Hour after Birth." The major aim of this study is to investigate the effects of pain medications administered during labor on the instinctive behavior patterns of newborn babies.

From the background of the study proposal: "The instinctive behavior patterns of normal, unmedicated newborn infants during the first hour after birth while in continuous skin-to-skin contact with their mothers have been well documented. It is known that maternal medications can be transferred to the fetus through placental circulation and the sedating effect to newborn infants of maternal narcotic administration prior to delivery has been well documented." Unmedicated newborns placed skin-to-skin with their mothers make their way through nine specific stages immediately following birth. These stages are: birth cry, relaxation, awakening, activity, rest, crawling, familiarization, suckling, and sleep [17]. In this paper, birth cry, relaxation, awakening, activity, and rest will be referred to as the "early stages" while crawling, familiarization, suckling, and sleep will be referred to as the "later stages." Healthy newborns move through these stages within the first two hours after birth. To date, no known studies have investigated the effects of intrapartum medications on these stages, while several studies have investigated the effects of intrapartum medications on breastfeeding [2, 7].

To evaluate the effects of administered medications, study investigators recorded on video the first hour of newborn activity following birth. Independent coders reviewed the videos and identified behaviors in the videos. From this data we will be able to investigate how much time each baby spent in each stage, whether there were interferences from hospital staff or family members, and correlations between this information and the medication data.

An unforeseen circumstance that arose during the data collection phase of the study was the unanticipated removal of several consenting mothers from the video

portion of the study due to medical complications with either the mother, the baby, or both. Of a total 67 mothers who were cleared to participate in the study, 20 could not be filmed following birth due to medical complications with either the mother or the baby. We chose to keep them in the study in order to investigate whether there was a difference between the two groups that caused these 20 mothers to "crash out," meaning they had complications that prevented filming. Henceforth these two groups of mothers will be referred to as "removed," describing the 20 mothers who could not be filmed and "non-removed," describing the 47 mothers who were filmed.

The formation of a Major Qualifying Project was proposed concerning the statistical analysis of the data collected for this study. Data collected include amounts and types of medications administered to consenting mothers during labor, 1 minute and 5 minute Apgar scores, time between first medication administration and birth in non-removed mothers, whether the baby experienced jaundice in non-removed mothers, weight loss between the times of birth and discharge from the hopsital for babies of non-removed mothers, feeding type for babies of non-removed mothers, video codes stating the amount of time spent in each stage and the amount of time crying, and codes describing each baby's first minute after birth. The medications we focused on in this project are Naropin, a pain medication, Pitocin, an induction medication (to hurry labor), and lactated ringers, which are bags of IV fluid.

The analysis performed on this data includes odds ratio analysis on removed vs non-removed mothers for various medication types and combinations. Additionally, correlation analysis was performed and we attempted to fit functions to the relationship between medications and time spent in many of the nine stages. Additionally, we performed many t tests to determine whether there are statistically significant differences between various sets of two distinct groups. This analysis was done to find evidence of causal relationships in the data.

Chapter 2

Questions to Investigate

This chapter presents the data types used in this paper and the questions we aimed to investigate. The data is broken up into categorical, integer, floating point, and boolean as each data type requires different tests be used. The questions are organized by what they're investigating overall and each question presents a rationale for its investigation.

2.1 Data Types

1. Categorical
   (a) Types of Medications (pain medications, induction medications, etc)
   (b) Feeding Type (breastfeeding or mix of formula and breast)
   (c) 9 Stages (birth cry, relaxation, awakening, activity, rest, crawling, familiarization, suckling, and sleep)
2. Integer
   (a) Numbers of Medications
   (b) Apgar Scores (designed to evaluate a baby's apprearance, pulse, grimace, activity, and respiration)
   (c) Weight Loss (percentage change between birth and discharge)
   (d) Number of Stages
3. Floating Point
   (a) Amounts of Medications (measured in milligrams or micrograms)
   (b) Length of Epidural (given in hours and minutes)
   (c) Approximate Labor Length (time from first medication administration to birth)

(d) Stage Lengths (given in minutes and seconds)

4. Boolean

   (a) Did the baby develop Jaundice?

   (b) Did the baby reach familliarization?

   (c) Did the baby reach suckling?

   (d) Did the baby cry for more than three minutes?

   (e) Did the baby spend more than 40 minutes crying and in the birth cry, awakening, activity, and rest stages combined?

   (f) Did the mother receive an epidural?

   (g) Did the mother receive Pitocin?

   (h) Did the baby go through 6 or more of the 9 stages?

2.2 Hypothesis Testing

We use hypothesis testing in this project to determine if there is a significant difference between two distinct groups. The test asks whether the data agrees more with a null hypothesis that there is no difference or an alternative hypothesis that one group's characteristic, such as amount of medication, is greater than another group's characteristic.

Section 2.2.1 examines questions that concern the difference between the removed and non-removed groups. Section 2.2.2 examines the relationships between medications received and the lengths of certain stages. Section 2.2.3 examines the relationship between certain medications and the approximate labor lengths of mothers in the study.

2.2.1 Inter-Group Proportions

1. Is a mother who received an epidural more likely to crash out than a mother who received no epidural?

   The purpose of this question is to determine if there is evidence that receiving an epidural can contribute to a mother's or baby's odds of crashing. The null hypothesis in this case is that there is no difference in the proportion of mothers who crashed out with an epidural and the proportion who crashed out without an epidural. The alternative hypothesis is that the proportion of mothers who crashed out with an epidural is higher than the proportion who crashed out without an epidural.

2. Is a mother who received Pitocin more likely to crash out than a mother who received no Pitocin?

   The purpose of this question is to determine if there is evidence that receiving Pitocin can contribute to a mother's or baby's odds of crashing. The null hypothesis in this case is that there is no difference in the proportion of mothers who crashed out who received Pitocin and the proportion who crashed out who did not receive Pitocin. The alternative hypothesis is that the proportion of mothers who crashed out with Pitocin is higher than the proportion who crashed out without Pitocin.

3. Is a mother who received Pitocin and an epidural more likely to crash out than a mother who received neither?

   The purpose of this question is to determine if there is evidence that the combination of Pitocin and an epidural can contribute to a mother's or baby's odds of crashing. The null hypothesis in this case is that there is no difference in the proportion of mothers who crashed out who received Pitocin and an epidural and the proportion of mothers who crashed out who received neither. The alternative hypothesis is that the proportion of mothers who crashed out who received Pitocin and an epidural is higher than the proportion who crashed out who received neither.

4. Is a mother who received Pitocin and an epidural more likely to crash out than a mother who received Pitocin with no epidural?

   The purpose of this question is to determine if there is evidence that Pitocin itself can contribute to a mother's or baby's odds of crashing. The null hypothesis in this case is that there is no difference in the proportion of mothers who crashed out who received Pitocin and an epidural and the proportion of mothers who crashed out who received only Pitocin. The alternative hypothesis is that the proportion of mothers who crashed out who received Pitocin and an epidural is higher than the proportion who crashed out who received only Pitocin.

5. Is a mother who received Pitocin and an epidural more likely to crash out than a mother who received an epidural with no Pitocin?

   The purpose of this question is to determine if there is evidence that receiving an epidural itself can contribute to a mother's or baby's chances of crashing. The null hypothesis in this case is that there is no difference in the proportion of mothers who crashed out who received Pitocin and an epidural and the proportion of mothers who crashed out who received only an epidural. The alternative hypothesis is that the proportion of mothers who crashed out who received Pitocin and an epidural is higher than the proportion who crashed out who received only an epidural

2.2.2 Medication vs. Stage Lengths

1. Is there a significant difference in the number of pain medications received among mothers whose babies spent greater than 40 minutes in early stages and those who did not?

   The purpose of this question is to determine if there is evidence that a higher number of pain medications affects the amount of time spent in the early stages. The null hypothesis in this case is that there is no difference in the number of medications received by mothers of babies who spent more than 40 minutes in the early stages and mothers of babies who spent less than 40 minutes in the early stage. The alternative hypothesis is that mothers of babies who spent more than 40 minutes in the early stages received more medications on average than mothers of babies who spent less than 40 minutes in the early stages.

2. Does receiving a higher number of pain medications increase or decrease the number of stages a baby goes through?

   The purpose of this question is to determine if there is evidence that babies who received more pain medications went through fewer stages than babies who received fewer pain medications. The line was drawn at 6 stages per the Healthy Children Project's request. The null hypothesis in this case is that there is no difference in the number of pain medications received by mothers of babies who went through 6 or more stages and the number of medications received by mothers of babies who went through 5 or fewer stages. The alternative hypothesis is that mothers of babies who went through 5 or fewer stages received more pain medications on average than mothers of babies who went through 6 or more stages.

3. Does receiving a higher number of pain medications increase or decrease the amount of crying a baby does in the first hour after birth?

   The purpose of this question is to determine if there is evidence that babies whose mothers received more pain medications cried less than babies whose mothers received fewer pain medications. The Healthy Children Project's hypothesis was that more pain medications would have a sedating effect on the baby and the line was drawn at 2 medications. The null hypothesis in this case is that there is no difference in the amount of crying done by babies who mothers received 3 or more pain medications and by babies whose mothers received 2 or fewer pain medications. The alternative hypothesis is that babies whose mothers received 3 or more medications cried less than babies whose mothers received 2 or fewer medications.

4. Does receiving Pitocin increase or decrease the number of stages a baby goes through?

The purpose of this question is to determine if there is evidence that Pitocin decreases the number of stages a baby goes through. The line was drawn at 6 stages per the Healthy Children Project's request. The null hypothesis in this case is that there is no difference in the amount of pitocin received by mothers of babies who went through 6 or more stages and by mothers of babies who went through 5 or fewer stages.

5. Does receiving Pitocin increase or decrease the amount of crying a baby does in the first hour after birth?

   The purpose of this question is to determine if there is evidence that Pitocin affects the amount of crying a baby does in the first hour. The null hypothesis in this case is that there is no difference in the amount of crying done by babies of mothers who received Pitocin and by babies of mothers who did not received Pitocin. The alternative hypothesis is that babies of mothers who received Pitocin cried more than babies of mothers who did not receive Pitocin.

2.2.3 Medication and Labor Times

1. What is the relationship between approximate labor length and whether the baby spent greater than 40 minutes in the early stages?

   The purpose of this question is to determine if there is evidence that longer labor lengths lead to more time in the early stages. The null hypothesis in this case is that there is no difference in approximate labor length among mothers of babies who spent more than 40 minutes in the early stages and mothers of babies who spent less than 40 minutes in the early stages. The alternative hypothesis is that mothers of babies who spent more than 40 minutes in the early stages had longer labor lengths than mothers of babies who spent less than 40 minutes in the early stages.

2. Does the average labor length for babies who reached the suckling stage differ from that of babies who did not reach the suckling stage?

   The purpose of this question is to determine if there is evidence that babies who reached suckling had a shorter average labor length than babies who did not reach suckling. The null hypothesis in this case is that there is no difference in the approximate labor length among mothers of babies who reached suckling and mothers of babies who did not reach suckling. The alternative hypothesis is that mothers of babies who reached suckling had shorter labor lengths than mothers of babies who did not reach suckling.

3. What is the relationship between epidural time and whether the baby reached suckling in the first hour after birth?

The purpose of this question is to determine if there is evidence that babies who reached suckling had a significantly shorter epidural time than babies who did not reach suckling. The null hypothesis in this case is that there is no difference between the epidural time among mothers of babies who reached suckling and among mothers of babies who did not reach suckling. The alternative hypothesis is that mothers of babies who reached suckling had shorter epidural times than mothers of babies who did not reach suckling.

4. What is the relationship between epidural time and whether the baby reached familiarization in the first hour after birth?

   The purpose of this question is to determine if there is evidence that babies who reached familiarization had a significantly shorter epidural time than babies who did not reach familiarization. The null hypothesis in this case is that there is no difference in epidural time among mothers of babies who reached familiarization and among mothers of babies who did not reach familiarization. The alternative hypothesis is that mothers of babies who reached familiarization had shorter epidural times than mothers of babies who did not reach familiarization.

5. What is the relationship between epidural time and whether the baby spent greater than 15 minutes in the rest stage?

   The purpose of this question is to determine if there is evidence that babies with longer epidurals spent more time in the rest stage. The null hypothesis in this case is that there is no difference in epidural time among mothers of babies who spent more than 15 minutes in the rest stage and among mothers of babies who spent less than 15 minutes in the rest stage. The alternative hypothesis is that mothers of babies who spent more than 15 minutes in the rest stage had longer epidural times than mothers of babies who spent less than 15 minutes in the rest stage.

6. What is the relationship between epidural time and whether the baby went through 6 or more stages?

   The purpose of this question is to determine if there is evidence that babies who went through 6 or more stages had shorter epidural times than babies who went through 5 or fewer stages. The null hypothesis in this case is that there is no difference in the epidural time among mothers of babies who went through 6 or more stages and among mothers of babies who went through 5 or fewer stages. The alternative hypothesis is that mothers of babies who went through 6 or more stages had shorter epidural times than mothers of babies who went through 5 or fewer stages.

7. What is the relationship between epidural time and whether the baby spent greater than 40 minutes crying and in the early stages?

The purpose of this question is to determine if there is evidence that babies with longer epidural times spent more time in the early stages than babies who had shorter epidural times. The null hypothesis is that there is no difference in the epidural time among mothers of babies who spent more than 40 minutes in the early stages and among mothers of babies who spent less than 40 minutes in the early stages. The alternative hypothesis is that mothers of babies who spent more than 40 minutes in the early stages had longer epidural times than mothers of babies who spent less than 40 minutes in the early stages.

2.2.4 Jaundice

1. How do different amounts of Naropin affect a baby's chances of developing jaundice?

   The purpose of this question is to determine if there is evidence that higher amounts of Naropin make a baby more likely to develop jaundice. The null hypothesis in this case is that there is no difference in the average amount of Naropin received by mothers of babies who developed jaundice and by mothers of babies who did not develop jaundice. The alternative hypothesis is that mothers of babies who developed jaundice received lower amounts of Naropin than mothers of babies who did not develop jaundice.

2. How does feeding type affect a baby's chances of developing jaundice?

   The purpose of this question is to determine if there is evidence that breastfed babies are less likely to develop jaundice. The null hypothesis in this case is that there is no difference in the incidence of jaundice among babies who breastfed and babies who were fed formula and brestfed. The alternative hypothesis is that babies who are formula fed are more likely to develop jaundice than babies who receive mixed feeding.

2.3 Descriptive Statistics

In this section we calculate statistics that describe the data collected. The correlations examine relationships between two variables when graphed. The differences in averages looks at whether there was a statistically significant difference in averages of medications administered between two groups.

2.3.1 Correlations

1. What is the relationship between baby weight loss and number of lactated ringers the mother received?

The purpose of this question is to determine if there is evidence that increased numbers of lactated ringers lead to increased weight loss following birth.

2. What is the relationship between epidural time and time the baby spent crying in the first hour?

   The purpose of this question is to determine if there is evidence that longer epidurals lead to either more or less crying than shorter epidurals.

2.3.2 Differences in averages

1. For Naropin, was the average amount received among mothers in the removed group significantly different than the average amount received among mothers in the non-removed group?

   The purpose of this question is to determine if there is evidence that more epidural medications increased the likelihood of a mother or baby crashing.

2. For pain medications, was the average number received among mothers in the removed group significantly different than the average number received among mothers in the non-removed group?

   The purpose of this question is to determine if there is evidence that more pain medications increased the likelihood of a mother or baby crashing.

Chapter 3

Statistical Testing Methods

There are a multitude of statistical tests and testing methods available to the modern statistician. This chapter discusses the tests used in this paper, their procedures, and the assumptions that make the results valid. This chapter also introduces the software used to conduct those tests and why those particular pieces of software were chosen for analysis.

3.1 Fisher's Exact Test

In our work, we use Fisher's exact test to determine whether there is a difference in proportions of two samples, where a proportion is the percentage of a sample with a certain characteristic [11, 16]. More generally, it can be used on more than two samples. The test is used to examine proportions of two separate groups that are affected by some property, usually a disease or medical condition. This test is used in this project when dealing with categorical data, each of which can be sorted into one of two specific categories. This test is valid when the sample sizes are smaller than 30 subjects per group, and can only be performed on data which can be sorted into a table. First, the "odds ratio" is calculated. The odds ratio is the ratio of proportions of the two samples and is discussed in Section 3.1.1. The test then calculates the significance level of that odds ratio to determine if it is statistically significant. It is called an exact test because the significance of the deviation from a chosen null hypothesis can be calculated precisely, a characteristic not shared by many statistical tests.

For example, the test can be used in this project to determine if there is a difference in the proportion of removed mothers who received an epidural and the proportion of non-removed mothers who received an epidural. For our example, a table for the data would be similar to the one given in Table 3.1. In this case, our null hypothesis is that there is no difference in proportions in the population and Fisher's exact test is telling us the significance of the difference in proportions of the sample. That is, the null hypothesis states that medication has no effect on a mother's chances

	Removed Mothers	Non-Removed Mothers	Row Totals
With Medication	a	b	$a+b$
Without Medication	c	d	$c+d$
Column Totals	$a+c$	$b+d$	$a+b+c+d$

Table 3.1: A sample 2x2 contingency table.

of crashing out, and our alternative hypothesis is that medication makes a mother more likely to crash out. For the purposes of this project, every instance of Fisher's exact test performed will be a one-sided test, not a two-sided test. In other words, our alternative hypothesis will always be that the odds ratio is greater than one, not solely that it does not equal one.

Table 3.1 is called a 2x2 contingency table. The properties name the rows and columns, with independent properties naming the rows and the possibly dependent properties naming the columns. Independent properties do not depend on any other property in the study and dependent properties change based on how the independent property changes. In this example, our categories are "with medication" and "without medication." The categories cannot overlap, that is, no data point should fall into two categories, and all data must be represented.

3.1.1 Odds Ratios

An odds ratio is a numerical measure of association between two possibly related properties of a set of data. It is a ratio of two proportions, measuring how much more likely a subject is to have a certain property given that they are in a certain of the two groups [13]. The odds ratio compares the incidence of a property in two separate samples and Fisher's Exact Test calculates the statistical significance of the odds ratio. Because since we are trying to find evidence of a causal relationship in this study, our two properties will be a definitely independent property and a possibly dependent property for all cases where we are using Fisher's Exact Test. An odds ratio equal to one indicates no association between the two properties, while a higher odds ratio is evidence for the hypothesis that the possibly dependent property is, in fact, dependent on the independent property. A higher odds ratio does not prove that there is a causal relationship because the possibly dependent property could still be caused by a third factor rather than being caused by the independent property. An odds ratio of less than one indicates that the property is more likely to occur in the second group rather than the first [1]. In our example, the odds ratio would be how much more likely a mother is to crash out if she receives medication than if she receives no medication.

When we calculate the odds ratio we can also calculate the confidence interval of that odds ratio. A confidence interval, associated with a specific confidence level,

is a range of values which, we hope, contains the actual population parameter. The confidence level, usually represented as a percentage, is a measure of how certain we are that the interval contains the true value.

3.1.2 Procedure

The first step in performing Fisher's exact test is to determine the sample odds ratio. To calculate the odds ratio, we refer to Figure 3.1 and follow this formula:

$$OR = \frac{a/b}{c/d} \tag{3.1.1}$$

Our next step in performing Fisher's exact test is calcuating the probability of obtaining a table with those specific values. This probability is given by the hypergeometric distribution as follows:

$$P = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{N}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!} \tag{3.1.2}$$

where $N = a + b + c + d$ is the total sample size. This is the probability of sampling the exact values a, b, c, and d given the row totals $a + b$ and $c + d$, the column total $a + c$, and the total sample size N. In our example, we know the number of medicated and non-medicated mothers, $a + b$ and $c + d$ respectively. We know who are the $a + b$ medicated mothers and the $c + d$ unmedicated mothers. There are $\binom{a+b}{a}\binom{c+d}{c}$ ways to select a women from the medicated group to be removed and c women from the unmedicated group to be removed. Overall, there are $\binom{N}{a+c}$ ways to choose $a+c$ women to be removed from the entire sample. Therefore, the probability of generating a table as in Table 3.1 is given by Equation (3.1.2).

The next step in performing Fisher's exact test is constructing tables with the same margin values (column and row totals) but with more extreme values for each space. In other words, we are constructing tables with higher odds ratios. We then calculate P as we did above for each of these tables and sum all of the values we have obtained. This sum is the p-value for the test.

3.1.3 p-values

A p-value is a value obtained in hypothesis testing and other types of significance testing. The p-value is the probability of obtaining a specific data set by chance from sampling if the null hypothesis is true. Lower p-values are evidence in favor of the alternative hypothesis and can often help us to reject the null hypothesis. In this study, our targeted significance level is 95%, so we will reject the null hypothesis in favor of the alternative hypothesis whenever our p-value is less than 0.05. When our p-value is greater than 0.05 we fail to reject the null hypothesis.

3.2 Welch's t Test

In Statistics, a t test is used to determine whether two independent samples, each from a different population, have significantly different means, indicating a difference in the actual population means. The choice of t test depends on whether the sample sizes are equal or unequal and whether the population variances are equal or unequal. The combination of mean and variance analysis is particularly important with a t test. Comparing the difference in means in meaningless if you do not account for the variance because if the difference is less than the "spread" of the data then it is probably not a significant difference. This test is specifically designed for use with samples having unequal sample sizes and from populations with possibly unequal variances [12]. In this project, we do not know the variances of our populations, and thus would have to use Welch's t test even if our sample sizes were equal, which they are not. Variance is a measure of spread in a population, which is why it is so important to know whether the variances are equal or unequal. Welch's t test calculates a t statistic which is then compared to a t distribution to determine the p-value of the test.

The null hypothesis for Welch's t test is that the two population means are equal. For the purposes of this paper, the alternative hypothesis will always be that a particular population mean is higher than the other. An important note about Welch's t test is a condition of the sample sizes. For sample sizes smaller than 15 subjects per group, the data must be approximately normally distributed. No such restriction is placed when the sample sizes are larger than 15 subjects per group.

3.2.1 Procedure

The t statistic is also called the test statistic and is a numerical summary of the sample calculated by considering the sample means, variances, and sizes. The t statistic for the Welch's t test is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{3.2.1}$$

where $\bar{X}_i$ is the average of sample i, s_i^2 is the variance of sample i, and N_i is the sample size of sample i for $i = 1, 2$. The next step in performing Welch's test is to calculate the degrees of freedom associated with the sampling. The degrees of freedom, ν, are a measure of how much the calculated value of the test statistic, t, can vary: the degrees of freedom are closely related to sample size and sample variance. To calculate ν we use the Welch-Satterthwaite equation which appears in equation (3.2.2) [15].

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2\nu_1} + \frac{s_2^4}{N_2^2\nu_2}} \tag{3.2.2}$$

where

$$\nu_i = N_i - 1 \tag{3.2.3}$$

is the degrees of freedom associated with s_i, the standard deviation of sample i [11]. Now that we have our t value and the degrees of freedom, we need only consult a t table to find the p-value for this test.

3.2.2 Student's t Distribution

In general, a probability distribution is an equation, sometimes represented graphically and sometimes represented as a table, that gives the probability (usually called the p-value) of a specific experimental outcome under the assumption that the null hypothesis is true. The distribution of the test statistic resulting from Welch's t test is approximately equal to an ordinary Student's t distribution. The Student's t distribution is a set of continuous probability distributions that result from estimating the average of one population if one sample is taken or the difference in averages of two populations when two samples are taken. These populations are assumed to be normally distributed, the sample sizes are assumed to be small, and the actual standard deviation of the populations is unknown.

The Student's t distribution varies based on the degrees of freedom of the test being performed. Figure 3.1 illustrates the t distribution with various degrees of freedom ν. Note that when $\nu = \infty$ the t distribution is exactly equal to the normal distribution. Along the y-axis we see the probability of any specific value of the experimental variable, the difference in means for our example. Along the x-axis we see the possible different values of $\bar{X}_1 - \bar{X}_2$, (on this graph labeled x). This is assuming we are considering two populations, and the values are given in terms of multiples of the pooled standard deviation of our two samples s_p[12], where

$$s_p = \frac{\sqrt{(N_1 - 1)\, s_1^2 + (N_2 - 1)\, s_2^2}}{N_1 + N_2 - 2}. \tag{3.2.4}$$

This data can also be represented as a table. Figure 3.2 is a sample t table for one- and two-tailed (the same as one- or two-sided) t tests. This figure can help in determining the approximate p-value of a t test. To use this table, one scans down the first column to find the degrees of freedom associated with the test, and then across that row to find between what two p-values the t statistic for the test lies. For example, say our number of degrees of freedom is 20 and $t = 1.9$. Then with our table we scan down to 20 in the left column, then across to see that 1.9 falls between 1.725 and 2.086. For a one sided test, this means that our p-value lies between 0.025 and 0.05, and for a two-sided test this means our p-value lies between 0.05 and 0.10. In practice, most statistical software will calculate the p-value exactly.

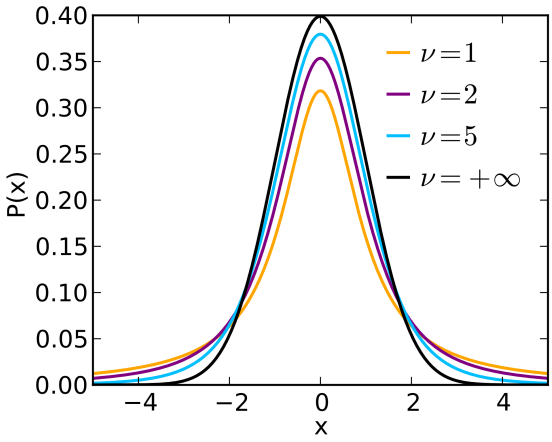

Figure 3.1: t distribution with varying degrees of freedom [5].

(1 tail)	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
(2 tail)	0.1	0.05	0.02	0.01	0.005	0.002	0.001
df							
1	6.3138	12.7065	31.8193	63.6551	127.3447	318.493	636.405
2	2.92	4.3026	6.9646	9.9247	14.0887	22.3276	31.5989
3	2.3534	3.1824	4.5407	5.8408	7.4534	10.2145	12.9242
4	2.1319	2.7764	3.747	4.6041	5.5967	7.1732	8.6103
5	2.015	2.5706	3.365	4.0322	4.7734	5.8934	6.8688
6	1.9432	2.4469	3.1426	3.7074	4.3168	5.2076	5.9589
7	1.8946	2.3646	2.998	3.4995	4.0294	4.7852	5.4079
8	1.8595	2.306	2.8965	3.3554	3.8325	4.5008	5.0414
9	1.8331	2.2621	2.8214	3.2498	3.8325	4.2969	4.7809
10	1.8124	2.2282	2.7638	3.1693	3.6896	4.2969	4.5869
11	1.7959	2.201	2.7181	3.0158	3.4966	4.0247	4.4369
12	1.7823	2.1788	2.681	3.0545	3.4284	3.9296	4.3178
13	1.7709	2.1604	2.6503	3.0123	3.3725	3.852	4.2208
14	1.7613	2.1448	2.6245	2.9768	3.3257	3.7874	4.1404
15	1.753	2.1314	2.6025	2.9467	3.286	3.7328	4.0728
16	1.7459	2.1199	2.5835	2.9208	3.252	3.6861	4.015
17	1.7396	2.1098	2.5669	2.8983	3.2224	3.6458	3.9651
18	1.7341	2.1009	2.5524	2.8784	3.1966	3.6105	3.9216
19	1.7291	2.093	2.5395	2.8609	3.1737	3.5794	3.8834
20	1.7247	2.086	2.528	2.854	3.1534	3.5518	3.8495

Figure 3.2: t table for use in conjunction with either a one- or two-tailed t test [6].

3.3 Correlation Testing

Correlation is a type of relationship between two variables which often indicates a dependent relationship. It is a measure of how much change in one variable indicates a change in another variable. Usually, correlation is measured with the Pearson Correlation Coefficient. This coefficient measures linear correlation in the relationships between the two variables.

3.3.1 Determining r

The Pearson Correlation Coefficient, written r, can be determined from a set of n paired observations. The value of r can be computed with the following formula:

$$r = \frac{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2}} \tag{3.3.1}$$

where N is the number of paired observations, X_i is the i^{th} observation of the first variable, $\bar{X}$ is the average of the first variable, Y_i is the i^{th} observation of the second variable, and $\bar{Y}$ is the average of the second variable. The value of r can be between 1 and -1, with values close to 1 indicating a strong positive linear relationship, values close to -1 indicating a strong negative linear relationship, and values near zero indicating little or no linear relationship between the two variables. Figure 3.3 illustrates several examples of data with specific r values.

3.4 Mann-Whitney U Test

The Mann-Whitney U test, also called the Mann-Whitney-Wilcoxon test or the Wilcoxon rank-sum test, is used to test whether two populations are the same. This test is credited with having a higher efficiency than the t test on populations with non-normal distributions. Originally developed for use with continuous data, the U test is just as valid on discrete data provided the test is formulated with specific assumptions in mind. The test is used with ordinal, ranked data and is used to determine differences between two independent samples. In other words, this data must be able to be ordered and ranked from lowest to highest.

There two assumptions that we make about the data when using the Mann-Whitney U test. The first is that the observations are independent of each other, meaning that one observation does not affect the next. The second assumption is that the distributions of the two populations are equal under the null hypothesis. This means that the probability of an observation from group 1 exceeding an observation from group 2 is equal to the probability of an observation from group 2

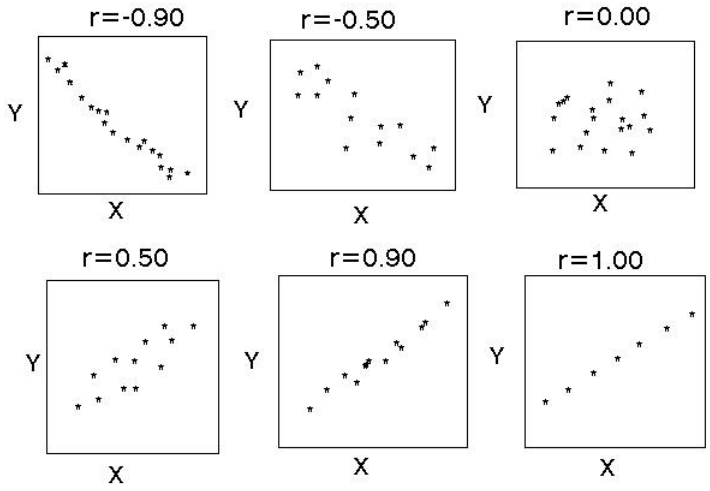

Figure 3.3: Scatterplots of data with various r values [10].

exceeding an observation from group 1. In other words, the probability distributions have the same shape and their means fall at the same place.

3.4.1 Procedure

The first step in performing the Mann-Whitney U test is determining the value of the test statistic U [9]. The value of U can be calculated with respect to both groups being tested. The calculation is conducted as follows:

$$U_i = N_1 N_2 + \frac{N_i(N_i + 1)}{2} - R_i \tag{3.4.1}$$

where R_i is the sum of the ranks in sample i and N_i is the sample size of sample i for $i = 1, 2$. This calculation is performed for each of the two groups and then the smaller U value is used in testing. For large samples, U follows an approximately normal distribution, and from U we can calculate a z value to use in consulting a table: z values are similar to t values, but are, in general, used with more ideal data. Calculated z values are used in conjunction with z tables which represent the values of a z distribution, much as with t values. This calculation is as follows:

$$z = \frac{U - \mu_U}{\sigma_U} \tag{3.4.2}$$

where μ_U and σ_U are, respectively, the mean and standard deviation of U. These values are calculated as follows:

$$\mu_U = \frac{N_1 N_2}{2} \tag{3.4.3}$$

$$\sigma_U = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}} \tag{3.4.4}$$

It is important to note that this calculation of σ_U is only valid for samples with very few ties and with no large tie bands. Ties occur when two values have the same rank, and tie bands occur when many values in order have the same rank. However, in practice, when using statistical software to calculate these parameters, an adjusted formula is used which can account for large numbers of ties or tie bands.

3.5 Statistical Software

3.5.1 R

R is a programming language and environment that allows users to perform statistical computing. It is an implementation of the S programming language and is widely used by statisticians, data miners, and data analysts. R was used to perform each incidence of Fisher's Exact Test in this paper. The program was chosen because it is open source and thus easily attainable and because it is widely used in industry.

3.5.2 Microsoft Excel

Excel is a spreadsheet program widely used for mathematical calculations and graphing. The industry standard for spreadsheets, Excel is used by high school students and professionals alike. Chosen for its wide usage and ability to use customizable formulae, Excel was used to perform the odds ratio analyses, t tests, and correlation analyses.

Chapter 4

Results

This chapter presents the results of our statistical tests. Each question has an explanation as to what test was used, what the hypotheses involved were, what the outcome of the test was, and whether the outcome was statistically significant.

4.1 Hypothesis Testing

4.1.1 Inter-Group Proportions

1. Is a mother who received an epidural more likely to crash out than a mother who received no epidural?

   When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.0784. The contingency table is shown in Table 4.1. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 3.167 with a 95% confidence interval for the true odds ratio of (0.800, 12.541). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that a mother who received an epidural is more likely to crash out than a mother who received no epidural.

2. Is a mother who received Pitocin more likely to crash out than a mother who

	Removed Mothers	Non-Removed Mothers	Row Totals
With Epidural	14	28	42
Without Epidural	3	19	22
Column Totals	17	47	64

Table 4.1: Contingency Table for Question 4.1.1-1

	Removed Mothers	Non-Removed Mothers	Row Totals
With Pitocin	12	26	38
Without Pitocin	5	21	26
Column Totals	17	47	64

Table 4.2: Contingency Table for Question 4.1.1-2

	Removed Mothers	Non-Removed Mothers	Row Totals
With Medication	10	17	27
Without Medication	1	10	11
Column Totals	11	27	38

Table 4.3: Contingency Table for Question 4.1.1-3

received no Pitocin?

When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.210. The contingency table is shown in Table 4.2. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 1.938 with a 95% confidence interval for the true odds ratio of (0.589, 6.380). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that a mother who received Pitocin is more likely to crash out than a mother who received no Pitocin.

3. Is a mother who received Pitocin and an epidural more likely to crash out than a mother who received neither? Analysis: Fisher's Exact Test, Odds Ratio.

   When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.0880. The contingency table is shown in Table 4.3. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 5.882 with a 95% confidence interval for the true odds ratio of (0.652, 53.039). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that a mother who received Pitocin and an epidural is more likely to crash out than a mother who received neither.

4. Is a mother who received Pitocin and an epidural more likely to crash out than a mother who received Pitocin with no epidural?

   When testing with Fisher's exact test against an alternative hypothesis that

	Removed Mothers	Non-Removed Mothers	Row Totals
With Epidural	10	17	27
Without Epidural	2	9	11
Column Totals	12	26	38

Table 4.4: Contingency table for Question 4.1.1-4. All mothers received Pitocin.

	Removed Mothers	Non-Removed Mothers	Row Totals
With Pitocin	10	17	27
Without Pitocin	4	11	15
Column Totals	14	28	42

Table 4.5: Contingency Table for Question 4.1.1-5. All mothers received an epidural.

the true odds ratio is less than 1 our p-value is 0.231. The contingency table is shown in Table 4.4. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 2.647 with a 95% confidence interval for the true odds ratio of (0.474, 14.780). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that a mother who received Pitocin and an epidural is more likely to crash out than a mother who received Pitocin with no epidural.

5. Is a mother who received Pitocin and an epidural more likely to crash out than a mother who received an epidural with no Pitocin?

   When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.371. The contingency table is shown in Table 4.5. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 1.618 with a 95% confidence interval for the true odds ratio of (0.405, 6.466). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that a mother who received Pitocin and an epidural is more likely to crash out than a mother who received an epidural with no Pitocin.

4.1.2 Medication vs. Stage Lengths

1. Is there a significant difference in the number of pain medications received among mothers whose babies spent greater than 40 minutes in the early stages and those who did not?

	≤ 5 Stages	≥ 6 Stages	Row Totals
≥ 3 Medications	11	13	24
≤ 2 Medication	4	16	20
Column Totals	15	29	44

Table 4.6: Contingency Table for Question 4.1.2-2

The average number of pain medications received by mothers of babies who spent greater than 40 minutes in the early stages was 2.379, while the average among mothers of babies who spent less than 40 minutes in the early stages was 1.400. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.0333, at the 95% confidence level we can reject the null hypothesis that there's no difference in the groups in favor of the alternative hypothesis that babies who spent greater than 40 minutes in the early stages received more pain medications than those who spent less than 40 minutes in the early stages. Therefore we can say that mothers whose babies spent more than 40 minutes in the early stages received more pain medications than mothers whose babies spent less than 40 minutes in the early stages.

2. Does receiving a higher number of pain medications increase or decrease the number of stages a baby goes through?

   When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.0683. The contingency table is shown in Table 4.6. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 3.385 with a 95% confidence interval for the true odds ratio of (0.870, 13.166). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that receiving a higher number of pain medications changes the number of stages a baby goes through.

3. Does receiving a higher number of pain medications increase or decrease the amount of crying a baby does in the first hour after birth?

   The average amount of time spent crying in the first hour after birth by babies whose mothers received two or fewer pain medications was 6 minutes, 31 seconds, while the average among babies whose mothers received three or more pain medications was 3 minutes, 19.9 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.0511, at the 95% confidence level we fail to reject the null hypothesis that there's no difference in the amount of crying between the two groups.

	≤ 5 Stages	≥ 6 Stages	Row Totals
With Pitocin	8	15	23
Without Pitocin	7	14	21
Column Totals	15	29	44

Table 4.7: Contingency Table for Question 4.1.2-4

Therefore we cannot say that a higher number of pain medications changes the amount of crying a baby does in the first hour after birth.

4. Does receiving Pitocin increase or decrease the number of stages a baby goes through?

   When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.586. The contingency table is shown in Table 4.7. We fail to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 1.067 with a 95% confidence interval for the true odds ratio of (0.306, 3.719). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that Pitocin changes the number of stages a baby goes through.

5. Does receiving Pitocin increase or decrease the amount of crying a baby does in the first hour after birth?

   The average amount of crying among babies whose mothers received Pitocin was 5 minutes, 19.9 seconds, while the average among babies whose mothers did not receive Pitocin was 4 minutes, 24 seconds. Welch's t test was performed to determine whether the difference in averages is statsitically significant. With a p-value of 0.311, at the 95% confidence level we fail to reject the null hypothesis that there's no difference in the average amount of crying between the two groups. Therefore we cannot say that Pitocin changes the amount of crying a baby does in the first hour after birth.

4.1.3 Medication and Labor Time

1. What is the relationship between approximate labor length and whether the baby spent greater than 40 minutes in the early stages?

   The average approximate labor length among mothers whose babies spent less than 40 minutes in the early stages was 11 hours, 28 minutes, and 30 seconds, while the average among mothers whose babies spent more than 40 minutes in the early stages was 12 hours, 5 minutes, and 31 seconds. Welch's t test was performed to determine whether the difference in averages is statisticallly

significant. With a p-value of 0.429, at the 95% confidence level we fail to reject the null hypothesis that there is no difference in approximate labor length between the two groups. Therefore we cannot say that there is any relationship between approximate labor length and whether the baby spent greater than 40 minutes in the early stages.

2. Does the average labor length for babies who reached the suckling stage differ from that of babies who did not reach the suckling stage?

   The average approximate labor length for mothers of babies who reached suckling was 6 hours, 56 minutes, and 28 seconds, while the average among mothers whose babies did not reach suckling was 15 hours, 11 minutes, and 26 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.00127, at the 95% confidence level we can reject the null hypothesis that there's no difference in the groups in favor of the alternative hypothesis that mothers whose babies reached suckling have shorter labor lengths than mothers whose babies did not reach suckling. Therefore we can say the average labor length for babies who reached the suckling stage was shorter than for babies who did not reach suckling.

3. What is the relationship between epidural time and whether the baby reached suckling in the first hour after birth?

   The average epidural time among mothers whose babies reached suckling was 2 hours, 43 minutes, and 25 seconds, while the average among mothers whose babies did not reach suckling was 7 hours, 3 minutes, and 24 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.00337, at the 95% confidence level we can reject the null hypothesis that there's no difference between the groups in favor of the alternative hypothesis that mothers of babies who reached suckling had shorter epidural times than mothers of babies who did not reach suckling. Therefore we can say that mothers of babies who reached suckling had shorter epidural times than mothers of babies who did not reach suckling.

4. What is the relationship between epidural time and whether the baby reached familiarization in the first hour after birth?

   The average epidural time among mothers whose babies reached the familiarization stage was 3 hours, 18 minutes, and 28 seconds, while the average for mothers whose babies did not reach familiarization was 7 hours, 49 minutes, and 40 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.0141, at the 95% confidence level we can reject the null hypothesis that there's no difference between the groups in favor of the alternative hypothesis that mothers of babies who reached familiarization had shorter epidural times than mothers of babies

who did not reach familiarization. Therefore we can say that mothers of babies who reached familiarization had shorter epidural times than mothers of babies who did not reach familiarization.

5. What is the relationship between epidural time and whether the baby spent greater than 15 minutes in the rest stage?

   The average epidural time among mothers whose babies spent more than 15 minutes in the rest stage was 5 hours, 13 minutes, and 48 seconds, while the average among mothers whose babies spent less than 15 minutes in the rest stages was 5 hours, 23 minutes, and 25 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.465, at the 95% confidence level we fail to reject the null hypothesis that there's no difference in the average between the two groups. Therefore we cannot say that there's a relationship between epidural time and whether the baby spent more than 15 minutes in the rest stage.

6. What is the relationship between epidural time and whether the baby went through 6 or more stages?

   The average epidural time among mothers whose babies went through 5 or fewer stages was 8 hours, 9 minutes, and 47 seconds while the average epidural time among mothers whose babies went through 6 or more stages was 3 hours, 31 minutes, and 52 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.101, at the 95% confidence interval we can refect the null hypothesis that there's no difference in the groups in favor of the alternative hypothesis that mothers of babies who went through 5 or fewer stages had longer epidural times than mothers whose babies went through 6 or more stages. Therefore we can say that mothers whose babies went through 5 or fewer stages had longer epidurals than mothers whose babies went through 6 or more stages.

7. What is the relationship between epidural time and whether the baby spent greater than 40 minutes crying and in the early stages?

   The average epidural time among mothers of babies who spent less than 40 minutes in the early stages was 2 hours, 26 minutes and 56 seconds, while the average among mothers of babies who spent more than 40 minutes in the early stages was 6 hours, 38 minutes, and 38 seconds. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.00329, at the 95% confidence level we can reject the null hypothesis that there's no difference in the groups in favor of the alternative hypothesis that mothers of babies who spent less than 40 minutes in the early stages had shorter epidural times than mothers of babies who spent more than 40 minutes in the early stages. Therefore we can say that mothers of babies who spent less

	Jaundice	No Jaundice	Row Totals
Mixed	6	7	13
Breastfed	9	18	27
Column Totals	15	25	40

Table 4.8: Contingency Table for Question 4.1.4-2

than 40 minutes in the early stages had shorter epidural times than mothers of babies who spent more than 40 minutes in the early stages.

4.1.4 Jaundice

1. How do different amounts of Naropin affect a baby's chances of developing jaundice?

   The average amount of Naropin received by mothers of babies who got jaundice was 46.52 mg, while the average among mothers of babies who did not get jaundice was 61.41 mg. Welch's t test was performed to determine whether the difference in averages is statistically significant. With a p-value of 0.264, at the 95% confidence level we fail to reject the null hypothesis that there is no difference in average amount of Naropin received between the two groups. Therefore we cannot say that Naropin affects a baby's chances of developing jaundice.

2. How does feeding type (formula & breastmilk versus breastmilk alone) affect a baby's chances of developing jaundice?

   When testing with Fisher's exact test against an alternative hypothesis that the true odds ratio is less than 1 our p-value is 0.329. The contingency table is shown in Table 4.8. We faill to reject the null hypothesis that the true odds ratio is 1 because our p-value is greater than 0.05. The sample odds ratio is 1.714 with a 95% confidence interval of (0.443, 6.629). The fact that this confidence interval contains 1 means we cannot say that the odds ratio is statistically significant. Therefore we cannot say that feeding type affects a baby's chances of developing jaundice.

4.2 Descriptive Statistics

4.2.1 Correlations

1. What is the relationship between baby weight loss and number of lactated ringers mother received?

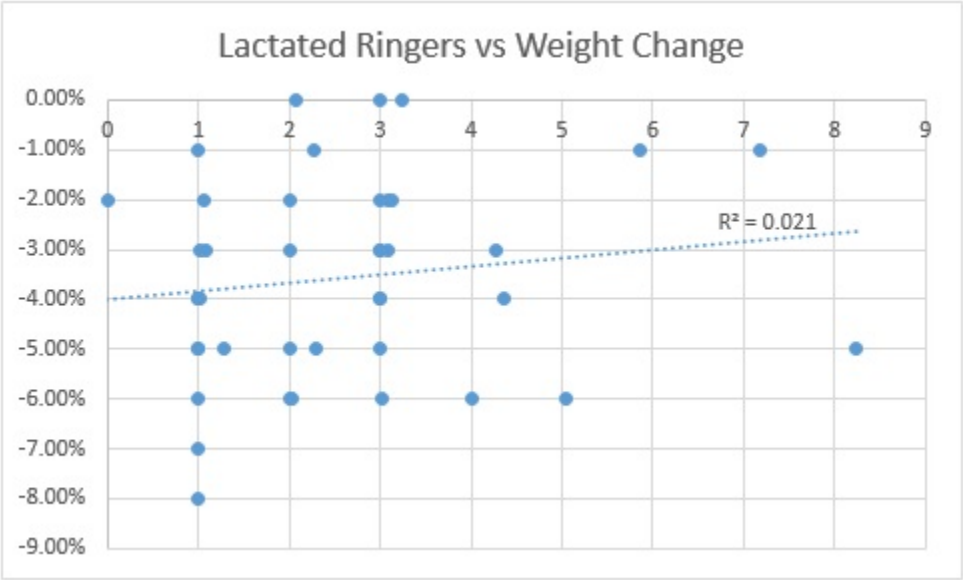

Figure 4.1: Scatterplot of lactated ringers received by mothers versus baby's weight change following birth

Based on the scatterplot of lactated ringers (Figure 4.1) values versus weight change values there is no relationship between the two. This hypothsis is substantiated by the linear regression performed with a Pearson's Correlation Coefficient of 0.145. Therefore we cannot say that there is a relationship between baby weight loss and the number of lactated ringers the mother received.

2. What is the relationship between epidural time and time the baby spent crying in the first hour?

   Based on the scatterplot of epidural times versus crying, one would expect that there is no relationship between the amount of epidural that a mother received and the time that her baby spent crying. This hypothesis is substantiated by the linear regression performed which gave a Pearson's Correlation Coefficient of -0.27 (Figure 4.2). When we considered only mothers who actually received an epidural (i.e. time greater than zero), the correlation coefficient is -0.161 (Figure 4.3). Therefore we cannot say that there is a relationship between epidural time and time the baby spent crying in the first hour after birth.

4.2.2 Differences in Averages

1. For Naropin, was the average amount received among mothers in the removed group significantly different than the average amount received among mothers in the non-removed group?

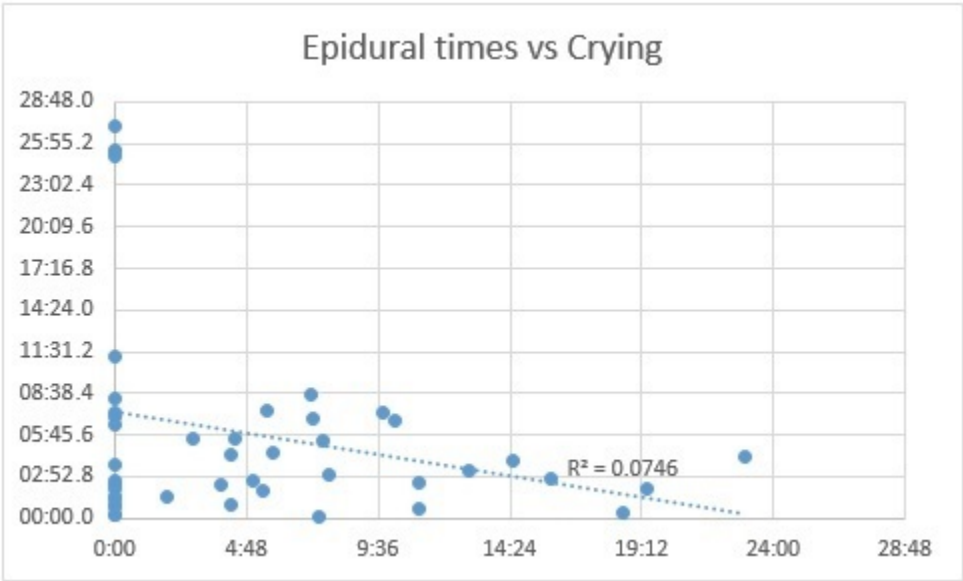

Figure 4.2: Scatterplot of length of time an epidural was used by a mother versus the amount of time the baby spent crying in the first hour.

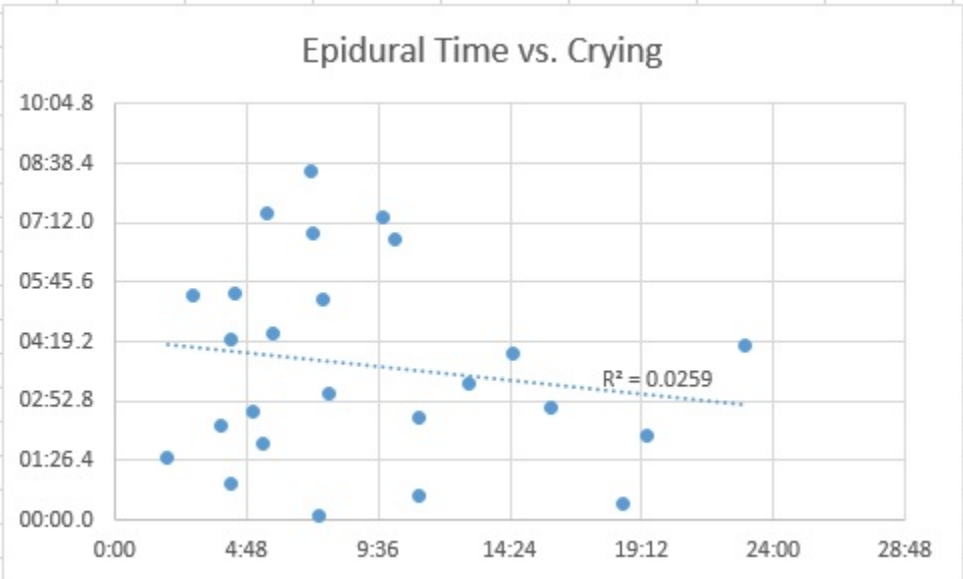

Figure 4.3: Scatterplot of length of time an epidural was used by a mother versus the amount of time the baby spent crying in the first hour after birth without zero epidural time.

The average amount of Naropin received in the removed group was 147.58 mg and in the non-removed group was 66.12 mg. This leaves us with a statistically significant difference in averages of 81.47, with Welch's t test giving us a p-value of 0.0135. The 95% confidence interval for the difference in averages is (12.96, 149.97). Therefore the average amount of Naropin received was higher among mothers in the removed group than among mothers in the non-removed group.

2. For Pitocin, was the average amount received among mothers in the removed group significantly different than the average amount received among mothers in the non-removed group?

   The average amount of Pitocin received in the removed group is 5602.65 mg and in the non-removed group was 3652.75 mg. This leaves us with a statistically insignificant difference of 1949.90, with Welch's t test giving us a p-value of 0.230. The 95% confidence interval for the difference in averages is (-3229.41, 7129.21). Therefore we cannot say that ther average amount of Pitocin received was different between the two groups.

Chapter 5

Conclusions

This chapter discusses the conclusions drawn from the tests performed on this data. Conclusions were drawn about the usefulness and applicability of each test and about the outcomes of the tests themselves. Recommendations for the future are presented at the end of this chapter.

5.1 Usefulness of Tests

5.1.1 Fisher's Exact Test

Fisher's Exact Test was particularly useful on this set of data. Many of the questions being asked were regarding data that could be meaningfully sorted into a 2x2 contingency table, making it perfect for Fisher's Exact Test. Additionally, this test allowed us to calculate precise p-values for our hypotheses, a characteristic not shared by Odds Ratio analysis on its own. Because of the small sample sizes associated with this data, Fisher's exact test was one of the only tests that could be validly used to answer the Healthy Children Project's questions.

5.1.2 Welch's t Test

Welch's t test was useful on this data because of the small sample sizes and non-normality displayed. This test allowed us to determine whether there were statistically significant differences in averages between two groups, something that can provide evidence of a causal relationship. The t test was also useful when we wanted to calculate confidence intervals regarding the difference in averages. For some of these questions we wanted to know what the possible range of the difference in averages was, and that is easy to calculate using the t test.

5.1.3 Correlation Testing

Correlation testing was not particularly useful with this data set. There were only two questions where it was the appropriate method of testing. Correlation testing can be useful when attempting to find patterns and the questions asked about this data was less about patterns and more about differences between two groups.The fact that our R values were low in the two instances does not mean that the correlation was meaningless in those cases, in fact it means that the test worked but there was not correlation.

5.1.4 Mann-Whitney U Test

The Mann-Whitney U Test did not turn out to be useful on this set of data. Among all of the data collected, some of the data was ordinal and could be ranked and that data would have worked well with analysis from a Mann-Whitney U Test. However, we did not end up analyzing that data under the scope of this project as the Healthy Children Project was more interested in other questions at this time.

5.2 Data Conclusions

5.2.1 Inter-Group Proportions

The results in Section 4.1.1 showed that there are no questions about inter-group proportions that yielded statistically significant results. This does not mean for sure that there was no difference between the removed mothers and the non-removed mothers, but rather means that our tests were inconclusive. As with all hypothesis testing, a high p-value does not mean we acccept the null hypotheses, but rather that we fail to reject it.

5.2.2 Medication vs. Stage Lengths

We had only one test in Section 4.1.2 with a statistically significant result. This test provides evidence for the hypothesis that the number of pain medications received by a mother contribute to the amount of time a baby spends in the early stages, with more medications probably contributing to more time in the early stages. The rest of the tests in Section 4.1.2 did not have statistically significant results, meaning we fail to reject the null hypotheses.

5.2.3 Medication and Labor Time

Several of the questions in Section 4.1.3 had statistically significant results after testing. There is meaningful evidence that longer labor lengths decrease the chance that

a baby will reach the suckling stage and that longer epidural times decrease a baby's chances of reaching suckling. There is also evidence that longer epidural times decrease a baby's chances of reaching familiarization and that they may decrease the number of stages a baby goes through overall. Longer epidural times may also lead to babies spending longer in the early stages than babies of mothers with shorter epidural times.

5.2.4 Jaundice

There were only two questions in Section 4.1.4. Neither of the questions had statistically significant results, so we are unable to draw any conclusions about the incidence of jaundice or the causes of jaundice.

5.2.5 Correlations

Neither of our two correlation tests revealed any relationships between are variables. This means the questions cannot be meaningfully answered after this analysis.

5.2.6 Differences in Averages

In Section 4.2.2 we only had two questions, one of which had a statistically significant testing result. We found that the average amount of Naropin received in the removed group was significantly larger than the average amount received in the non-removed group. The same test was conducted for Pitocin with no statistically significant result.

5.3 Recommendation

Given the fact that many of these tests were inconclusive, more research, preferably with larger sample sizes, is needed. However, with our participants we have uncovered significant evidence that epidural pain medications may have a negative effect on a baby's first hour after birth. There's evidence that these medications reduce the number of instinctive stages the baby goes through and that they reduce the baby's chances of reaching critical later stages such as familiarization and suckling. Since we know that all of the babies and mothers were healthy prior to entering labor, and because we know that the vast majority of unmedicated babies go through all nine stages, it's reasonable to hypothesize that these medications are having a negative effect.

Bibliography

[1] Deng. "How to interpret odds ratios that are smaller than 1?." On Biostatistics and Clinical Trials. 12 February 2012. 8 Jan 2014. <http://onbiostatistics.blogspot.com/2012/02/how-to-interpret-odds-ratios-that-are.html>.

[2] Dozier, AM, CR Howard, EA Brownell, RN Wissler, JC Glantz, SR Ternullo, KN Thevenet-Morrison, CK Childs, and RA Lawrence. "Labor epidural annesthesia, obstetric factors and breastfeeding cessation." *Maternal Child Health Journal*. 17.4 (2013): 689-698. Web. 29 Jan 2014. <http://www.ncbi.nlm.nih.gov/pubmed/22696104>.

[3] "Fisher's Exact Test for Count Data." *R Documentation*. Web. 29 Sept 2013. <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/fisher.test.html>.

[4] healthychildren.cc

[5] http://en.wikipedia.org/wiki/File:Student_t_pdf.svg

[6] http://homes.ieu.edu.tr/hcetinkaya/t_table.jpg

[7] Jordan, S, S Emery, C Bradshaw, A Watkins, and W Friswell. "The impact of intrapartum analgesia on infant feeding." *British Journal of Obstetrics and Gynaecology*. 112.7 (2005): 927-934. Web. 29 Jan 2014. <http://www.ncbi.nlm.nih.gov/pubmed/15957994>.

[8] Lane, David. "Difference Between Means." *Online Statistics Education: An Interactive Multimedia Course of Study*. Rice University. Web. 13 Oct 2013. <http://onlinestatbook.com/lms/estimation/differencemeans.html>.

[9] Mann, H. B., and D. R. Whitney. "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other." *The Annals of Mathematical Statistics* 18.1 (1947): 1-164. Web. 29 Jan 2014. <http://projecteuclid.org/euclid.aoms/1177730491>.

[10] http://ordination.okstate.edu/corr.jpg

[11] Petruccelli, Joseph, Balgobin Nandram, and Minghui Chen. *Applied Statistics for Engineers and Scientists.* 1st ed. Upper Saddle River, New Jersey: Prentice Hall, 1999, Print.

[12] Ruxton, Graeme. "The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test." *Behavioral Ecology.* 17.4 (2006): 688-690. Web. 10 Oct 2013. <http://beheco.oxfordjournals.org/content/17/4/688.full>.

[13] Szumilas, Magdalena. "Explaining Odds Ratios" *Journal of the Canadian Academy of Child and Adolescent Psychiatry.* 19.3 (2010): 227-229. Web. 15 Oct 2013. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/>.

[14] Telke, Susan. "Confidence Intervals of Odds Ratio and Relative Risk." Web. 17 Oct 2013. <http://www.biostat.umn.edu/~susant/Fall10ph6414/Lesson14complete.pdf>.

[15] "Two-Sample T-Tests Allowing Unequal Variance (Enter Difference)." NCSS Statistical Software. NCSS, LLC. Web. 28 Jan 2014. <http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/PASS/Two-Sample_T-Tests_Allowing_Unequal_Variance-Enter_Difference.pdf>.

[16] Weisstein, Eric. "Fisher's Exact Test." MathWorld-A Wolfram Web Resource Web. 25 Sept 2013. <http://mathworld.wolfram.com/FishersExactTest.html>.

[17] Widstrom, AM, V Wahlberg, AS Matthiesen, P Eneroth, K Uvnas-Moberg, S Werner, and J Winberg. "Short-term effects of early suckling and touch of the nipple on maternal behaviour." *Early Human Development.* 21.3 (1990): 153-163. Web. 30 Jan 2014. <http://www.ncbi.nlm.nih.gov/pubmed/2311552>.

[18] Yau, Chi. "Matrix." *R Tutorial: An R Introduction to Statistics.* Web. 29 Sept 2013. <http://www.r-tutor.com/r-introduction/matrix>

Chapter 6

Appendices

6.1 R code

This is the R code used for all instances of Fisher's Exact Test in this paper. The hashtags indicate comments in the code.

```
# Fisher's Exact Test for question 4.1.1-1
A=matrix(c(14,28,3,19),ncol=2,byrow=TRUE)
fisher.test(A)

#Fisher's Exact Test for question 4.1.1-2
B=matrix(c(12,26,5,21),ncol=2,byrow=TRUE)
fisher.test(B)

#Fisher's Exact Test for question 4.1.1-3
C=matrix(c(10,17,1,10),ncol=2,byrow=TRUE)
fisher.test(C)

#Fisher's Exact Test for question 4.1.1-4
D=matrix(c(10,17,2,9),ncol=2,byrow=TRUE)
fisher.test(D)

#Fisher's Exact Test for question 4.1.1-5
E=matrix(c(10,17,4,11),ncol=2,byrow=TRUE)
fisher.test(E)

#Fisher's Exact Test for question 4.1.2-2
F=matrix(c(11,13,4,16),ncol=2,byrow=TRUE)
fisher.test(F)
```

```
#Fisher's Exact Test for question 4.1.2-4
G=matrix(c(8,15,7,14),ncol=2,byrow=TRUE)
fisher.test(G)

#Fisher's Exact Test for question 4.1.4-2
H=matrix(c(6,7,9,18),ncol=2,byrow=TRUE)
fisher.test(H)
```

6.2 Excel

This is what a spreadsheet of medical record data looks like. We cannot include the actual spreadsheet because it would violate HIPAA, but this is a close approximation to the actual Excel sheet.

Patient #	Naropin	Pitocin	Apgar	Epidural	Stages	Time from first med	Jaundice	Baby weight	Feeding	Patient #	Suckling	Familiarization
1	130.2	0	9	0:00	4	0:27	no	-7.00%	BF	1	No	No
2	69.88	0	6	0:00	6	2:54	no	-5.00%	BF	2	Yes	Yes
3	0	345	6	0:00	6	12:20	no	-3.00%	BF	3	No	Yes
4	20	10848	6	11:04	8	23:38	no	-2.00%	Mixed	4	Yes	Yes
5	82.63	0	8	0:00	6	2:39	yes	-5.00%	Mixed	5	Yes	No
6	115	0	9	0:00	8	1:48	no	-1.00%	BF	6	Yes	Yes
7	25	0	9	7:48	7	11:58	no	-4.00%	BF	7	Yes	Yes
8	39.7	1747	9	0:00	9	4:28	yes	-3.00%	Mixed	8	Yes	Yes
9	0	21	8	5:34	5	7:16	no	-2.00%	Mixed	9	No	No
10	0	24712	9	10:11	3	26:12	no	-5.00%	Mixed	10	No	No
11	0	0	7	0:00	4	15:34	no	-5.00%	BF	11	No	No
12	109.45	1350	8	7:28	6	10:03	no	0.00%	BF	12	Yes	Yes
13	0	0	8	7:37	5	10:50	yes	-2.00%	Mixed	13	No	No
14	0	7050	8	11:04	7	24:33	no	-4.00%	BF	14	No	Yes
15	92.4	0	8	0:00	7	0:34	yes	-6.00%	BF	15	No	Yes
16	0	0	7	0:00	7	0:38	no	-3.00%	BF	16	No	Yes
17	73.5	0	8	0:00	6	1:07	no	-8.00%	BF	17	Yes	Yes
18	129.4	709	9	8:11	7	10:38	no	-6.00%	BF	18	No	No
19	0	17250	7	19:24	4	26:12	no	-1.00%	Mixed	19	No	No
20	35.66	0	9	4:16	7	8:34	yes	-2.00%	BF	20	Yes	Yes
21	77.55	1606	8	5:02	6	14:35	no	-2.00%	Mixed	21	No	No
22	68.15	23694	7	0:00	6	53:55	no	-1.00%	BF	22	No	Yes
23	0	0	8	7:13	7	9:00	no	0.00%	BF	23	Yes	Yes
24	0	0	8	1:54	5	11:58	no	-4.00%	BF	24	No	No
25	0	109	8	0:00	6	22:24	no	-5.00%	BF	25	No	Yes
26	55.8	4753	7	18:32	4	18:32	no	0.00%	Mixed	26	No	No
27	390.4	5662	8	11:11	5	19:47	yes	-3.00%	BF	27	No	No
28	55	1018	8	0:00	7	5:21	no	-2.00%	Mixed	28	Yes	Yes
29	64.8	5476	8	0:00	5	13:46	no	-1.00%	Mixed	29	Yes	Yes
30	0	0	9	15:55	4	17:16	yes	-3.00%	BF	30	No	No